UNIVERSIDAD CARLOS III DE MADRID

documentos de trabajo

Documento de Trabajo 03-01 (01)

Serie de Estadística y Econometría

Enero 2003

Departamento de Estadística y Econometría

Universidad Carlos III de Madrid

Calle Madrid, 126

28903 Getafe (España)

Fax (34) 91 624-98-49

Comportamiento asintótico del correlograma residual

Andrés Ubierna y Santiago Velilla *

Resumen

En este trabajo se formalizan algunos resultados relativos al comportamiento asintótico del correlograma residual que, en la literatura de series temporales, aparecen recurrentemente y son de gran utilidad para la obtención y demostración de algunas propuestas de interés. Como ejemplo, se incluye una derivación del estadístico de Box-Pierce (1970).

Palabras Clave:

Modelos ARMA; Correlación residual; Convergencia.

*Ubierna, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, C/Madrid, 126, 28903 Getafe (Madrid). España, e-mail: aubierna@est-econ.uc3m.es, Tel: +34 916249890. Velilla, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, e-mail: savece@est-econ.uc3m.es. Los autores se han beneficiado parcialmente del proyecto de investigación BEC2000-0167 del plan nacional de I+D+I.

1. INTRODUCCIÓN

En lo que sigue, se supone que

$$\phi(B)(X_t - \mu) = \theta(B)\varepsilon_t \ , \quad t = 0, \ \pm 1, ..., \tag{1}$$

es un proceso $ARMA(p\,,q)$ causal e invertible de media μ, donde B es el operador de retardos $BX_t = X_{t-1}$ y $\{\varepsilon_t : t = 0, \pm 1, ..\}$ es una secuencia de errores i.i.d. con media cero, varianza σ^2 y momento de cuarto orden $E[\varepsilon_t^4]$ finito. Además, $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ ... \ -\phi_p z^p$ y $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \ ... \ +\theta_q z^q$ son polinomios de grados p y q respectivamente, en los que $\phi_p \neq 0$ y $\theta_q \neq 0$. Se supone también que los polinomios $\phi(z)$ y $\theta(z)$ no tienen raíces comunes y que éstas se sitúan fuera del círculo unidad. En concreto, se supone que si r_1, ..., r_p son las raíces de $\phi(z)$ y s_1, ..., s_q son las de $\theta(z)$, se cumple $r_i \neq s_j$, $|r_i|$, $|s_j| > 1$ y también, para $i \neq j$, $r_i \neq r_j$ y $s_i \neq s_j$. Se define además el parámetro $P = \max(p, q)$.

Es convientente, por motivos de claridad en la derivación de estos resultados, comenzar por el caso $AR(p)$ generalizando más adelante al caso $ARMA(p, q)$.

2. CASO $AR(P)$

En el caso en que $\boldsymbol{\theta}(z) = 1$ y, por tanto, $P = p$, (1) se convierte en

$$\boldsymbol{\phi}(B)(X_t - \mu) = \varepsilon_t, \quad t = 0, \pm 1, ...,. \tag{2}$$

Para este modelo podemos definir las funciones $\varepsilon_t(\boldsymbol{\phi}, \mu)$ de acuerdo a la siguiente expresión,

$$\varepsilon_t(\boldsymbol{\phi}, \mu) = (X_t - \mu) - \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu) \ .$$

Estas coinciden con los ruidos ε_t para $t > p$. En la práctica se consideran únicamente las funciones $\varepsilon_t(\boldsymbol{\phi}, \mu)$ para $t > p$ con el fin de evitar la dependencia de las condiciones iniciales, $X_t - \mu = 0 \ t \leq 0$. Claramente, a partir de la expresión para $\varepsilon_t(\boldsymbol{\phi}, \mu)$, una vez hayamos estimado $\boldsymbol{\phi}$ y μ tendremos de un modo natural expresiones para los residuales del modelo estimado

Dada una muestra $\mathbf{X}_n = (X_1, ..., X_n)'$, la media μ se estima mediante la media muestral $\hat{\mu} = \overline{X}_n = (X_1 + ... + X_n)/n$. Es conocido que $\sqrt{n}(\overline{X}_n - \mu)$ es asintóticamente $N(0, a)$, donde $a > 0$ (ver, por ejemplo, Brockwell y Davis, 1991, Cap. 7). Por otro lado, los estimadores de $\boldsymbol{\phi}$ que se consideran son los estimadores de tipo Yule - Walker (Brockwell y Davis, 1991, Cap. 8),

$$\widehat{\boldsymbol{\phi}} = \arg\min_{\phi} \sum_{t>p}^{n} [\varepsilon_t(\boldsymbol{\phi}, \overline{X}_n)]^2 = \arg\min_{\phi} \sum_{t>p}^{n} [(X_t - \overline{X}_n) - \sum_{j=1}^{p} \phi_j (X_{t-j} - \overline{X}_n)]^2.$$

Estos estimadores satisfacen un sistema de ecuaciones lineales de la forma $\sum_{t>p}^{n} \varepsilon_t(\boldsymbol{\phi}, \overline{X}_n)(X_{t-j} - \overline{X}_n) = 0$, $j = 1, ..., p$, y tienen, por tanto, una expresión explícita fácil de determinar. Además, $\sqrt{n}(\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \xrightarrow{D} N_p[\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\phi})]$,donde $\mathbf{I}(\boldsymbol{\phi})$ es una matriz definida positiva de $p \times p$ (Brockwell y Davis, 1991, Cap. 8).

Los residuales $\widehat{\varepsilon}_t$ se obtienen con la expresión $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\phi}}, \overline{X}_n)$, y, para $t > p$, se calculan numéricamente, de manera natural, mediante la expresión $\widehat{\varepsilon}_t = (X_t - \overline{X}_n) - \sum_{j=1}^{p} \widehat{\phi}_j (X_{t-j} - \overline{X}_n)$.

2.1 Momentos muestrales de los errores y de los residuales

Es interesante estudiar las propiedades asintóticas de los dos primeros momentos de los $\widehat{\varepsilon}_t$, $t > p$. Para ello resulta útil relacionar los errores con los residuales a traves de la identidad,

$$\widehat{\varepsilon}_t = \varepsilon_t - \sum_{j=1}^{p} (\widehat{\phi}_j - \phi_j)(X_{t-j} - \mu) + Q_n \ , \quad t > p \ , \tag{3}$$

donde $Q_n = -\widehat{\boldsymbol{\phi}}(1)(\overline{X}_n - \mu)$ y $\widehat{\boldsymbol{\phi}}(1) = 1 - \sum_{j=1}^{p} \widehat{\phi}_j$. El miembro derecho de (3), que no es factible porque depende de cantidades desconocidas, nos ayudará a estudiar propiedades asintóticas de los dos primeros momentos muestrales de los $\widehat{\varepsilon}_t$, $t > p$.

Introducimos ahora los dos primeros momentos muestrales de los residuales y de los errores para estudiar posteriormente la conexión que existe entre ellos.

	Residuales	Errores
Media muestral	$\widehat{p}_n = \sum_{t>p}^{n} \widehat{\varepsilon}_t / n$	$p_n = \sum_{t=1}^{n} \varepsilon_t / n$
Función de autocovarianza	$\widehat{g}_k = \sum_{t>p}^{n-k} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t+k} / n$	$g_k = \sum_{t=1}^{n-k} \varepsilon_t \varepsilon_{t+k} / n$

(4)

donde $0 \leq k \leq n-1$.

Dado que el proceso $AR(p)$ se ha supuesto estacionario, podemos, a partir de la serie $\boldsymbol{\phi}^{-1}(z) = \sum_{r=0}^{\infty} h_r z^r$, presentar los coeficientes h_r, que cumplen la acotación $|h_r| \leq AB^r \leq A$,donde $A > 0$ y $0 < B < 1$. Con el convenio de notación $h_r = 0$ para $r < 0$ podemos obtener, a partir de los coeficientes h_r, la representación causal de nuestro proceso $AR(p)$ (ver, por ejemplo, Brockwell y Davis, 1991, Cap. 3):

$$X_t - \mu = \sum_{r=0}^{\infty} h_r \varepsilon_{t-r}. \tag{5}$$

Teorema 1 *En el caso $AR(p)$, si los errores $\{\varepsilon_t : t \in Z\}$ del modelo (2) son iid con media cero, varianza σ^2 y, además, verifican $E(\varepsilon_t^4) < +\infty$, se tiene:*

(a)

$$\widehat{p}_n = p_n + O_p(n^{-1/2}) \ ; \tag{6}$$

(b)

$$\widehat{g}_k = g_k - \sigma^2 \sum_{j=1}^{p} (\widehat{\phi}_j - \phi_j) h_{k-j} + O_p(n^{-1}) \ . \tag{7}$$

Demostración del Teorema 1. **(a)** Usando (3), se obtiene

$$\widehat{p}_n = \frac{1}{n} \sum_{t>p}^{n} \widehat{\varepsilon}_t = p_n + H_n \ ,$$

donde $H_n = -\sum_{t=1}^{p} \varepsilon_t / n + (n-p) Q_n / n - \sum_{j=1}^{p} (\widehat{\phi}_j - \phi_j) Q_j$ y $Q_j = n^{-1} \sum_{t=p+1}^{n} (X_{t-j} - \mu)$, $j = 1, ..., p$. Por tanto, basta demostrar que todos los sumandos que componen el término H_n son $O_p(n^{-1/2})$.

Por una parte, $\sqrt{n}\,|\sum_{t=1}^{p}\varepsilon_t/n| \leq p/\sqrt{n}\max_{1\leq t\leq p}|\varepsilon_t| = O_p(1)$, y $\sqrt{n}\,|(n-p)\,Q_n/n| = (n-p)\,/n\left|\widehat{\phi}(1)\right|\sqrt{n}\left|\overline{X}_n-\mu\right| = O_p(1)$. Por otro lado, para acotar la suma $\sum_{j=1}^{p}(\widehat{\phi}_j - \phi_j)Q_j$, se aplica, para cada $j = 1, ..., p$, el teorema 7.1.2 en Brockwell y Davies (1991, p. 219) al proceso lineal $Y_{t,j} = X_{p+t-j}$ para obtener $Q_j = (n-p)/n(1/(n-p)\sum_{t=p+1}^{n}X_{t-j}-\mu) = (n-p)/n(1/(n-p)\sum_{t=1}^{n-p}Y_{t,j}-\mu) = O_p(n^{-1/2})$ .Por tanto,

$$\sqrt{n}\mid\sum_{j=1}^{p}(\widehat{\phi}_j-\phi_j)Q_j\mid\leq\sqrt{n}\max_{1\leq j\leq p}\mid Q_j\mid\sum_{j=1}^{p}\mid\widehat{\phi}_j-\phi_j\mid= O_p(1)\ .$$

Observar que todos los términos de H_n son $O_p(n^{-1})$ salvo el segundo, que se debe a la consideración de una media μ posiblemente diferente de cero.

(b) Usando de nuevo la expresión (3) tenemos que $\widehat{\varepsilon}_t = \varepsilon_t - \sum_{j=1}^{p}(\widehat{\phi}_j-\phi_j)(X_{t-j}-\mu) + Q_n$ y $\widehat{\varepsilon}_{t+k} = \varepsilon_{t+k} - \sum_{j=1}^{p}(\widehat{\phi}_j-\phi_j)(X_{t+k-j}-\mu)+Q_n$. Realizando los productos cruzados de estas expresiones junto con la expresión causal del modelo $AR(p)$ (5), podemos llegar a la siguiente identidad,

$$\widehat{g}_k = g_k - \sigma^2\sum_{j=1}^{p}(\widehat{\phi}_j-\phi_j)h_{k-j} + U_{k,n}, \tag{8}$$

donde el resto $U_{k,n}$ puede expresarse como la siguiente combinación lineal finita

$$U_{k,n} = \frac{1}{n}\sum_{j=1}^{p^2+4p+3}\alpha_j X_{k,j}, \tag{9}$$

donde los pares $(\alpha_j, X_{k,j})$ son de la forma

α_j	$X_{k,j}$
1	a_k
$\sqrt{n}(\widehat{\phi}_j-\phi_j)$	$b_{k,j}/c_{k,j}$
$\sqrt{n}\widehat{\phi}(1)(\overline{X}_n-\mu)$	d_k
$n(\widehat{\phi}_j-\phi_j)(n^{-1}\sum_{t>p}^{n}\varepsilon_t^2-\sigma^2)$	h_{k-j}
$n(\widehat{\phi}_i-\phi_i)(\widehat{\phi}_j-\phi_j)$	$e_{k,i,j}$
$n[\widehat{\phi}(1)]^2(\overline{X}_n-\mu)^2$	f_k
$(\widehat{\phi}_j-\phi_j)$	$h_{k-j}(\sum_{t=n-k+1}^{n}\varepsilon_t^2)$

. (10)

Enunciamos ahora el siguiente lema que se demuestra en el apéndice y facilita el final de la demostración.

Lema 1 *En las condiciones del teorema 1, si definimos* $a_k = -\sum_{t=1}^{p} \varepsilon_t \varepsilon_{t+k}$, $b_{k,j} = -\sum_{r=0}^{\infty} h_r \left(\sum_{t>p}^{n-k} \varepsilon_{t-j-r} \varepsilon_{t+k} \right) / \sqrt{n}$, $c_{j,k} = -\sum_{r \neq k-j}^{\infty} h_r \left(\sum_{t>p}^{n-k} \varepsilon_t \varepsilon_{t+k-j-r} \right) / \sqrt{n}$, $d_k = -\sum_{t>p}^{n-k} (\varepsilon_t \varepsilon_{t+k}) / \sqrt{n}$, $e_{k,i,j} = \sum_{t>p}^{n-k} (X_{t-i} - \mu)(X_{t+k-j} - \mu)/n$, $f_k = (n-k-p)/n$ *y* $l_{k,j} = h_{k-j} \sum_{t=n-k+1}^{n} \varepsilon_t^2$*; se puede afirmar que existe una constante universal* $c > 0$*, independiente de los subíndices* i, j, k *y* n*, tal que*

$$\sup_{i,j,k,n} \left\{ E(a_k^2), E(b_{k,j}^2), E(c_{k,j}^2), E(d_k^2), h_{k-j}^2, E(e_{k,i,j}^2), f_k^2, E(l_{k,j}^2) \right\} \leq c \ . \tag{11}$$

Por tanto, el término $U_{k,n}$ de (9) es una suma finita de términos α_j con coeficientes $X_{k,j}$ como los definidos en el lema 1, de acuerdo con el cual, $E\left[X_{k,j}^2\right] \leq c$ y la demostración de este apartado se limitaría, de este modo, a estudiar el orden probabilístico de los términos α_j. En concreto, conocemos que $\sqrt{n}(\widehat{\phi}_j - \phi_j)$ y $\sqrt{n}(\overline{X}_n - \mu)$ son asintóticamente normales, al igual que lo es, por el teorema central del límite, $n^{1/2}(\sum_{t>p}^{n} \varepsilon_t^2/n - \sigma^2)$. Con lo que es evidente que todos los α_j son $O_p(1)$ u $o_p(1)$. Así, podemos concluir que $nU_{k,n} = O_p(1)$, o, lo que es o mismo, que $U_{k,n}$ es $O_p(n^{-1})$.

2.2 Comportamiento asintótico del correlograma residual

Como aplicación, sea $m = m_n$ de modo que se cumpla $m/n \to 0$. La expresión (8) se puede escribir de forma matricial en la forma

$$\widehat{\mathbf{G}}_m = \mathbf{G}_m - \sigma^2 \, \mathbf{X}_m (\widehat{\boldsymbol{\phi}} - \boldsymbol{\phi}) + \mathbf{U}_{m,n} \ , \tag{12}$$

donde $\widehat{\mathbf{G}}_m = (\widehat{g}_1, ..., \widehat{g}_m)'$, $\mathbf{G}_m = (g_1, ..., g_m)'$, $\mathbf{U}_{m,n} = (U_{1,n},\ U_{2,n},\ ...,\ U_{m,n})'$ y $\mathbf{X}_m$ es una matriz de orden $m \times p$ con estructura

$$\mathbf{X}_m = \left(\begin{array}{cccc} 1 & 0 & \cdots & 0 \\ h_1 & 1 & \cdots & \vdots \\ \vdots & \vdots & & \vdots \\ h_{p-1} & h_{p-2} & \cdots & 1 \\ \hline h_p & h_{p-1} & \cdots & h_1 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ h_{m-1} & h_{m-2} & \cdots & h_{m-p} \end{array} \right) .$$

En los pasos siguientes vamos a necesitar del siguiente resultado, que se demuestra en el apéndice.

Proposición 1 (Condiciones de Ortogonalidad) *La matriz* $\mathbf{X}_m$ *y el vector de covarianzas residuales* $\widehat{\mathbf{G}}_m$ *siempre que* $m = m_n \underset{n\to\infty}{\rightarrow} \infty$, *cumplen*

$$\mathbf{X}_m' \widehat{\mathbf{G}}_m = O_p(\frac{1}{n}) \ . \tag{13}$$

Como aplicación, sea $m = m_n = \sqrt{n}$ y $\mathbf{H}_m = \mathbf{X}_m(\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'$ la matriz de proyección ortogonal sobre el espacio de columnas $C(\mathbf{X}_m)$. Se tiene

$$\frac{\sqrt{n}}{\sigma^2}\widehat{\mathbf{G}}_m = \frac{\sqrt{n}}{\sigma^2}[\mathbf{H}_m\widehat{\mathbf{G}}_m + (\mathbf{I}_m - \mathbf{H}_m)\widehat{\mathbf{G}}_m] = (\mathbf{I}_m - \mathbf{H}_m)(\frac{\sqrt{n}}{\sigma^2}\mathbf{G}_m) + \mathbf{Z}_{m,n} \ , \tag{14}$$

donde $\mathbf{Z}_{m,n} = \sqrt{n}\mathbf{H}_m\widehat{\mathbf{G}}_m/\sigma^2 + (\mathbf{I}_m - \mathbf{H}_m)\sqrt{n}\mathbf{U}_{m,n}/\sigma^2$. Teniendo en cuenta las condicións de ortogonalidad, se tiene la desigualdad $\|\mathbf{Z}_{m,n}\| \leq \sqrt{n}\left\|\mathbf{H}_m\widehat{\mathbf{G}}_m\right\|/\sigma^2 +$

$\sqrt{n}\left\|\mathbf{U}_{m,n}\right\|/\sigma^2 = \sqrt{n}\left\|(\mathbf{X}_m'\mathbf{X}_m)^{-1/2}\mathbf{X}_m'\widehat{\mathbf{G}}_m\right\|/\sigma^2 + \sqrt{n}\left\|\mathbf{U}_{m,n}\right\|/\sigma^2 \leq$ $\leq \left\|(\mathbf{X}_m'\mathbf{X}_m)^{-1/2}\right\|\sqrt{n}\left\|\mathbf{X}_m'\widehat{\mathbf{G}}_m\right\|/\sigma^2 + \sqrt{n}\left\|\mathbf{U}_{m,n}\right\|/\sigma^2 = O_p(n^{-1/2}) + O_p(m^{1/2}n^{-1/2}) =$ $O_p(n^{-1/4})$, y, en particular $\left\|\mathbf{Z}_{m,n}\right\| \overset{P}{\to} 0$. En definitiva, la expresión (14) se puede aproximar

$$\frac{\sqrt{n}}{\sigma^2}\,\widehat{\mathbf{G}}_m \overset{D}{\cong} (\mathbf{I}_m - \mathbf{H}_m)(\frac{\sqrt{n}}{\sigma^2}\,\mathbf{G}_m) \overset{D}{\cong} (\mathbf{I}_m - \mathbf{H}_m)\mathbf{N}_m(\mathbf{0},\mathbf{I}_m)\ .$$

Teniendo en cuenta que $\widehat{g}_0 \overset{P}{\to} \sigma^2$, se concluye que

$$\sqrt{n}\,\widehat{\mathbf{R}}_m \overset{D}{\cong} (\mathbf{I}_m - \mathbf{H}_m)\mathbf{N}_m(\mathbf{0},\mathbf{I}_m)\ , \tag{15}$$

siendo $\widehat{\mathbf{R}}_m = (\widehat{r}_1, ..., \widehat{r}_m)'$ un vector de $m \times 1$ formado por las m primeras correlaciones residuales $\widehat{r}_j = \widehat{g}_j/\widehat{g}_0$. De esta expresión (15) puede obtenerse la aproximación en distribución para el estadístico de bondad de ajuste de Box y Pierce (1970)

$$\widehat{Q}_n = n\sum_{k=1}^{m}\widehat{r}_k^2 = n\widehat{\mathbf{R}}_m'\widehat{\mathbf{R}}_m \overset{D}{\cong} \mathbf{U}_m'(\mathbf{I}_m - \mathbf{H}_m)\mathbf{U}_m \sim \chi^2_{m-p}\ . \tag{16}$$

3. CASO $ARMA(P,Q)$

En el caso $ARMA(p,q)$, el modelo al que nos referimos es el introducido en (1). En este caso, las funciones $\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)$, $t = 0, \pm 1, ...$, se calcularán a partir de la siguiente recursión

$$\begin{aligned}
&\varepsilon_1(\boldsymbol{\phi},\boldsymbol{\theta},\mu) = (X_1 - \mu)\\
&\varepsilon_2(\boldsymbol{\phi},\boldsymbol{\theta},\mu) = (X_2 - \mu) - \phi_1\,(X_1 - \mu) - \theta_1\varepsilon_1(\boldsymbol{\phi},\boldsymbol{\theta},\mu)\\
&...\\
&\varepsilon_n(\boldsymbol{\phi},\boldsymbol{\theta},\mu) = \phi\,(B)\,(X_n - \mu) - \theta_1\varepsilon_{n-1}(\boldsymbol{\phi},\boldsymbol{\theta},\mu) - ... - \theta_q\varepsilon_{n-q}(\boldsymbol{\phi},\boldsymbol{\theta},\mu)
\end{aligned}$$

Esta recursión puede resumirse en la expresión $\phi(B)(X_t - \mu) = \theta(B)\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)$, junto con las condiciones iniciales $X_t - \mu \equiv 0 \equiv \varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)$ para $t \leq 0$. Por tanto,

$$\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu) = \sum_{j=0}^{t-1}\pi_j(X_{t-j} - \mu), \tag{17}$$

donde los $\{\pi_j : j \geq 0\}$ son los coeficientes de la serie $\pi(z) = \phi(z)/\theta(z)$. La suma (17) depende únicamente de $\{X_1 - \mu,\ ...,\ X_t - \mu\}$. La relación de las funciones $\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$ con los errores ε_t se deduce de la condición de invertibilidad del proceso, que permite escribir $\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu) = \varepsilon_t - q_t$, donde $q_t = \sum_{j=t}^{\infty} \pi_j (X_{t-j} - \mu)$. En la práctica, se consideran únicamente las funciones $\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$ asociadas a valores de $t > P$ para, en cierto sentido, evitar la dependencia de las condiciones iniciales $X_t - \mu \equiv 0 \equiv \varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$, $t \leq 0$.

Dada la muestra $\mathbf{X}_n = (X_1,\ ...,\ X_n)'$, μ se estima, como en el caso autoregresivo, mediante la media muestral $\overline{X}_n = (X_1 +\ ...\ + X_n)/n$. Los estimadores de $(\boldsymbol{\phi}, \boldsymbol{\theta})$ que se consideran son los de mínimos cuadrados, que se obtienen minimizando la función objetivo

$$(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}) = \arg \min_{(\boldsymbol{\phi}, \boldsymbol{\theta})} u_n(\boldsymbol{\phi}, \boldsymbol{\theta}) = \arg \min_{(\boldsymbol{\phi}, \boldsymbol{\theta})} \sum_{t>P}^{n} [\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \overline{X}_n)]^2 \ . \tag{18}$$

El problema numérico asociado es no lineal y, por tanto, la determinación de los estimadores $(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}})$ requiere la utilización de un algortimo iterativo. Finalmente, se sabe que,

$$\sqrt{n} \begin{pmatrix} \widehat{\boldsymbol{\phi}} - \boldsymbol{\phi} \\ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \end{pmatrix} \xrightarrow{D} N_{p+q}[\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\phi}, \boldsymbol{\theta})] \ , \tag{19}$$

donde $\mathbf{I}(\boldsymbol{\phi}\ ,\ \boldsymbol{\theta})$, la *matriz de información*, es una matriz simétrica de $(p+q) \times (p+q)$ definida positiva (Brockwell y Davis, 1991, Cap. 8).

Los residuales se definen a partir de las funciones $\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)$ introducidas en (17), en notación obvia como

$$\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \overline{X}_n) = \tag{20}$$

$$= \widehat{\boldsymbol{\theta}}^{-1}(B) \widehat{\boldsymbol{\phi}}(B)(X_t - \overline{X}_n) = \widehat{\boldsymbol{\pi}}(B)(X_t - \overline{X}_n) = \sum_{j=0}^{t-1} \widehat{\pi}_j (X_{t-j} - \overline{X}_n) \ ,$$

donde $\widehat{\boldsymbol{\phi}}$ y $\widehat{\boldsymbol{\theta}}$ son los estimadores de mínimos cuadrados condicionales. Los residuales,

$\widehat{\varepsilon}_t$, se calculan numéricamente usando la recursión,

$$\widehat{\varepsilon}_t = (X_t - \overline{X}_n) - \sum_{i=1}^{p} \widehat{\phi}_i (X_{t-i} - \overline{X}_n) - \sum_{j=1}^{q} \widehat{\theta}_j \widehat{\varepsilon}_{t-j}, \quad t = 1, ..., n, \tag{21}$$

en combinación con las condiciones iniciales $X_t - \overline{X}_n \equiv 0 \equiv \widehat{\varepsilon}_t$ para $t \leq 0$. En la práctica, sólo se consideran residuales $\widehat{\varepsilon}_t$ para $t > P$.

3.1 Momentos muestrales de los errores y de los residuales

Al igual que en el caso $AR(p)$, ahora estamos interesados en obtener propiedades asintóticas para los dos primeros momentos muestrales de los residuales. Para ello, consideramos un desarrollo de Taylor para $\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\phi}}, \widehat{\boldsymbol{\theta}}, \overline{X}_n)$ que nos permite expresar los residuales para $t > P$ como

$$\widehat{\varepsilon}_t = \varepsilon_t - \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) \boldsymbol{\theta}^{-1}(B)(X_{t-i} - \mu) - \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_j) \boldsymbol{\theta}^{-1}(B) \varepsilon_{t-j} + \widehat{H}_t \ , \tag{22}$$

donde $\widehat{H}_t = -q_t + \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) R_{t-i} + \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_j) S_{t-j} + F_t(\overline{X}_n - \mu) + \widehat{Z}_t$. Los detalles del desarrollo de Taylor así como expresiones explícitas para q_t, R_t, S_t, F_t y $\widehat{Z}_t$ pueden encontrarse en el apéndice. El miembro derecho de (22), que no es factible porque depende de cantidades desconocidas, sirve para estudiar propiedades de los dos primeros momentos muestrales de los $\widehat{\varepsilon}_t$, $t > P$.

Definimos como en (4) los dos primeros momentos muestrales de residuales y errores. La formulación es coincidente debido a la inclusión del parámetro $P = \max(p, q)$ que en ella aparece. Introducimos una nueva serie, $\boldsymbol{\eta}(z) = \boldsymbol{\theta}(z)/\boldsymbol{\phi}(z) = \sum_{r=0}^{\infty} \eta_r z^r$, que permite obtener a partir de sus coeficientes la representación causal del modelo $ARMA(p, q)$, que tendrá ahora la expresión

$$X_t - \mu = \sum_{r=0}^{\infty} \eta_r \varepsilon_{t-r}. \tag{23}$$

Recordemos que $\{h_r : r \geq 0\}$ son los coeficientes de la serie $h(z) = \phi^{-1}(z) = \sum_{j=0}^{\infty} h_r z^r$ y $\{l_r : r \geq 0\}$ son los coeficientes de la serie $l(z) = \theta^{-1}(z) = \sum_{r=0}^{\infty} l_r z^r$, lo que unido al convenio de notación $h_r = l_r = 0$ para $r < 0$ nos permite enunciar el siguiente teorema:

Teorema 2 *En el caso $ARMA(p,q)$, si los errores $\{\varepsilon_t : t \in \mathbb{Z}\}$ del modelo son i.i.d. con media cero, varianza σ^2 y, además, verifican, $E[\varepsilon_t^4] < +\infty$, se tiene:*

(a)

$$\widehat{p}_n = p_n + O_p(n^{-1/2}) \ ;$$

(b)

$$\widehat{g}_k = g_k - \sigma^2 \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) h_{k-i} - \sigma^2 \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_j) l_{k-j} + O_p(n^{-1}) \ .$$

Demostración.(a)

$$\widehat{p}_n = \frac{1}{n} \sum_{t=\max(p,q)+1}^{n} \widehat{\varepsilon}_t = p_n + H_n,$$

donde $H_n = -\sum_{t=1}^{\max(p,q)} \varepsilon_t/n - \sum_{t=\max(p,q)}^{n} \sum_{j=t}^{\infty} \pi_j (X_{t-j} - \mu)/n + \sum_{t=\max(p,q)}^{n} \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) D_{t,i}/n + \sum_{t=\max(p,q)}^{n} \sum_{i=1}^{p} (\widehat{\theta}_i - \theta_i) E_{t,j}/n + \sum_{t=\max(p,q)}^{n} F_t(\overline{X}_n - \mu)/n + \sum_{t=\max(p,q)}^{n} \widehat{Z}_t/n$. Se trata entonces de demostrar que cada uno de los sumandos que componen H_n son $O_p(n^{-1/2})$.

En concreto,

$$D_{t,i} \ = \ \partial\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)/\partial\phi_i = -\sum_{r=0}^{t-i-1} l_r (X_{t-i-r} - \mu) \ i = 1, ..., p, \tag{24}$$

$$E_{t,j} \ = \ \partial\varepsilon_t(\boldsymbol{\phi}, \boldsymbol{\theta}, \mu)/\partial\theta_j = -\sum_{r=0}^{t-j-1} l_r \varepsilon_{t-j-r}(\boldsymbol{\beta}, \mu) \ j = 1, ..., q, \tag{25}$$

mientras que las expresiones explícitas para F_t y $\widehat{Z}_t$ pueden verse en el apéndice. El primer sumando de H_n es, de hecho, $O_p(n^{-1})$ porque depende de un número finito de variables. En lo que respecta a los otros sumandos, tenemos por ejemplo para el término en $D_{t,j}$: $n^{-1} \sum_{t=\max(p,q)}^{n} \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) D_{t,i} = n^{-1} \sum_{t=\max(p,q)}^{n} \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i)[\sum_{r=0}^{t-i-1} l_r (X_{t-i-r} - \mu)]$. Si se definen $q_{i,t} = \sum_{r=0}^{t-i-1} l_r (X_{t-i-r} - \mu)$, este término se puede escribir, después de multiplicar por $\sqrt{n}$

$$\sum_{i=1}^{p} \sqrt{n}(\widehat{\phi}_i - \phi_i)[\frac{1}{n} \sum_{t=\max(p,q)}^{n} q_{i,t}] \ . \tag{26}$$

Por consiguiente, basta demostrar que, para $i = 1, .., p$, se cumple $n^{-1} \sum_{t=\max(p,q)}^{n} q_{i,t} = O_p(1)$. Teniendo en cuenta que $\sum_{r=1}^{\infty} |l_r| < +\infty$ se tiene

$$E[|q_{i,t}|] \leq c = [\gamma(0)]^{1/2} \sum_{r=1}^{\infty} |l_r| < +\infty \ , \tag{27}$$

de modo que $E[\left| n^{-1} \sum_{t=\max(p,q)}^{n} q_{i,t} \right|] \leq c$, lo que es suficiente para garantizar $n^{-1} \sum_{t=\max(p,q)}^{n} q_{i,t} = O_p(1)$. Actuando de un modo similar con el resto de sumandos, terminaríamos con esta parte de la demostración.

(b) Utilizando la expresión (22) y la representación causal del modelo, podemos llegar a la ecuación siguiente,

$$\widehat{g}_k = g_k(\boldsymbol{\beta}, \mu) - \sigma^2 \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) h_{k-i} - \sigma^2 \sum_{i=1}^{q} (\widehat{\theta}_i - \theta_i) l_{k-i} + U_{k,n} \ , \tag{28}$$

donde ahora el resto $U_{k,n}$ es más complejo que antes ya que resulta de multiplicar de modo cruzado y promediar los sumandos de las expansiones para $\widehat{\varepsilon}_t$ y $\widehat{\varepsilon}_{t+k}$. De hecho, según (22) tendremos

$$\widehat{\varepsilon}_t = \varepsilon_t - \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) \boldsymbol{\theta}^{-1}(B)(X_{t-i} - \mu) - \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_j) \boldsymbol{\theta}^{-1}(B) \varepsilon_{t-j} + \widehat{H}_t \ ,$$

$$\widehat{\varepsilon}_{t+k} = \varepsilon_{t+k} - \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i) \boldsymbol{\theta}^{-1}(B)(X_{t+k-i} - \mu) - \sum_{j=1}^{q} (\widehat{\theta}_j - \theta_j) \boldsymbol{\theta}^{-1}(B) \varepsilon_{t+k-j} + \widehat{H}_{t+k} \ .$$

Por ejemplo, desarrollando uno de estos sumandos, tendremos lo siguiente, $\sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i)[\frac{1}{n} \sum_{t>P}^{n-k} \varepsilon_t \theta^{-1}(B)(X_{t+k-i} - \mu)] = \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i)[\sum_{r=0}^{\infty} \sum_{s=0}^{\infty} l_r \eta_s$ $(\frac{1}{n} \sum_{t>P}^{n-k} \varepsilon_t \varepsilon_{t+(k-i)-(r+s)})]$, que puede descomponerse en los siguientes dos sumandos,

$$\sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i)[\sum_{(r+s) \neq (k-i)} l_r \eta_s (\frac{1}{n} \sum_{t>P}^{n-k} \varepsilon_t \varepsilon_{t+(k-i)-(r+s)})],$$

y

$$\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)[\sum_{(r+s)=(k-i)} l_r\eta_s(\frac{1}{n}\sum_{t>P}^{n-k}\varepsilon_t^2)].$$

El primero de ellos puede escribirse como sigue

$$\frac{1}{n}\sum_{i=1}^{p}\sqrt{n}(\widehat{\phi}_i-\phi_i)c_{A,k,i}+\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}(\frac{1}{n}\sum_{t>P}^{n-k}\varepsilon_t^2)], \tag{29}$$

donde $c_{A,k,i} = \sum_{(r+s)\neq(k-i)} l_r\eta_s(\frac{1}{\sqrt{n}}\sum_{t>P}^{n-k}\varepsilon_t\varepsilon_{t+(k-i)-(r+s)})$. Por su parte, el segundo sumando se puede escribir

$$\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}(\frac{1}{n}\sum_{t>P}^{n-k}\varepsilon_t^2) =$$

$$=\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}(\frac{1}{n}\sum_{t>P}^{n}\varepsilon_t^2)-\frac{1}{n}\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}(\sum_{t=n-k+1}^{n}\varepsilon_t^2)= \tag{30}$$

$$=\sigma^2\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}+\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)(\frac{1}{n}\sum_{t>P}^{n}\varepsilon_t^2-\sigma^2)h_{k-i}-\frac{1}{n}\sum_{i=1}^{p}(\widehat{\phi}_i-\phi_i)h_{k-i}(\sum_{t=n-k+1}^{n}\varepsilon_t^2).$$

De este término, por tanto, obtenemos el segundo término de (28). El resto de términos de (30) así como el término (29) pasan a formar parte del resto.

Actuando de modo análogo con todos los productos cruzados resultantes de multiplicar $\widehat{\varepsilon}_t$ y $\widehat{\varepsilon}_{t+k}$, llegamos a expresar el resto como una combinación lineal finita del tipo

$$nU_{k,n}=\sum_{m=1}^{Q}\beta_m Y_{k,m},$$

donde, $Q=[3+3(p+q)+(p+q)^2]^2+3(p+q)$, $\beta_m=O_p(1)$ u $o_p(1)$ y $E[Y_{k,m}^2]\leq c$, donde c es una constante positiva universal. La demostración de esta última propiedad es más compleja que la del lema 2 del caso autoregresivo, aunque utiliza argumentos similares. ■

3.2 Comportamiento asintótico del correlograma residual

Como aplicación de este resultado, del mismo modo que en el caso $AR(p)$, sea $m = m_n$, con $m_n/n \to 0$ la expresión (28) puede formularse de un modo matricial como sigue

$$\widehat{\mathbf{G}}_m = \mathbf{G}_m - \sigma^2 \mathbf{X}_m(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \mathbf{U}_{m,n} \ , \tag{31}$$

donde $\widehat{\boldsymbol{\beta}} = (\widehat{\phi}_1, ..., \widehat{\phi}_p, \widehat{\theta}_1, ..., \widehat{\theta}_q)'$, $\boldsymbol{\beta} = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)'$ y $\mathbf{U}_{m,n} = (U_{1,n}, U_{2,n}, ..., U_{m,n})'$. Además, la matriz $\mathbf{X}_m$, en el caso en que $p < q$, tiene la siguiente estructura

$$\mathbf{X}_m = \left(\begin{array}{cccc|cccc} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ h_1 & 1 & \cdots & \vdots & l_1 & 1 & \cdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ h_{p-1} & h_{p-2} & \cdots & 1 & \vdots & \vdots & & \vdots \\ \cline{1-4} h_p & h_{p-1} & \cdots & h_1 & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & l_{q-1} & l_{q-2} & \cdots & 1 \\ \cline{5-8} \vdots & \vdots & & \vdots & l_q & l_{q-1} & \cdots & l_1 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ h_{m-1} & h_{m-2} & \cdots & h_{m-p} & l_{m-1} & l_{m-2} & \cdots & l_{m-q} \end{array} \right) . \tag{32}$$

Las estructura de (31), caso $ARMA(p, q)$, y (12), caso $AR(p)$, son similares. Necesitamos, al igual que en el caso $AR(p)$ la siguiente proposición, que se demuestra en el apéndice.

Proposición 2 (Condiciones de Ortogonalidad) *La matriz* $\mathbf{X}_m$ *y el vector de covarianzas residuales* $\widehat{\mathbf{G}}_m$*, siempre que* $m = m_n \underset{n\to\infty}{\to} \infty$*, cumplen*

$$\mathbf{X}_m' \widehat{\mathbf{G}}_m = O_p(\frac{1}{n}) \ . \tag{33}$$

Este resultado, permite obtener para el caso $ARMA(p,q)$ una aproximación en distribución análoga a (15), donde la construcción de $\mathbf{H}_m$ sigue siendo como antes, $\mathbf{H}_m = \mathbf{X}_m(\mathbf{X}_m'\mathbf{X}_m)^{-1}\mathbf{X}_m'$, con la particularidad de que para el caso general, la matriz $\mathbf{X}_m$ es como en (32). Al igual que en caso $AR(p)$, podríamos derivar análogamente el estadístico de Box-Pierce (1970):

$$\widehat{Q}_n = n\sum_{k=1}^{m} \widehat{r}_k^2 \overset{D}{\cong} \mathbf{U}_m'(\mathbf{I}_m - \mathbf{H}_m)\mathbf{U}_m \sim \chi^2_{m-(p+q)} \ . \tag{34}$$

3.3 El enfoque de Box y Pierce

Otra expresión alternativa para (22), propuesta por Box y Pierce (1970), consiste en considerar el proceso autoregresivo $\{Y_t\}$ de orden $p+q$

$$\boldsymbol{\phi}(B)\boldsymbol{\theta}(B)(Y_t - \mu) = \boldsymbol{\theta}(B)\boldsymbol{\phi}(B)(Y_t - \mu) = \varepsilon_t \ , \tag{35}$$

que, se relaciona con el proceso original $\{X_t\}$ mediante la identidad $\boldsymbol{\theta}^2(B)(Y_t - \mu) = X_t - \mu$. Introduciendo esta información en (22), se tiene $\widehat{\varepsilon}_t = \varepsilon_t - \sum_{i=1}^{p}(\widehat{\phi}_i - \phi_i)\boldsymbol{\theta}(B)(Y_{t-i} - \mu) - \sum_{j=1}^{q}(\widehat{\theta}_j - \theta_j)\boldsymbol{\phi}(B)(Y_{t-j} - \mu) + \widehat{H}_t$, lo que permite expresar los residuales como

$$\widehat{\varepsilon}_t = \varepsilon_t - \sum_{i=1}^{p+q} \widehat{\lambda}_i(Y_{t-i} - \mu) + \widehat{H}_t \ , \quad t > P \ , \tag{36}$$

donde, en (36), $\widehat{\lambda}_i = \sum_{r=\max(0,i-p)}^{\min(q,i-1)} \theta_r(\widehat{\phi}_{i-r} - \phi_{i-r}) + \sum_{s=\max(0,i-q)}^{\min(p,i-1)} \phi_s(\widehat{\theta}_{i-s} - \theta_{i-s}) = O_p(n^{-1/2})$,siendo $\phi_0 = \theta_0 = 1$. Las expresiones [(22)/(36)] generalizan adecuadamente la relación (3), es decir, si $\theta(z) = 1$, [(22)/(36)] se reducen a (3). Observar también la semejanza estructural entre las expresiones (36) del caso general y (3) del caso autoregresivo.

Si ahora introducimos, a partir de los polinomios autorregresivo $(\phi(B))$ y de medias móviles $(\theta(B))$ la siguiente serie, con sus correspondientes coeficientes,

$$\mathbf{a}(z) = [\boldsymbol{\phi}(z)\boldsymbol{\theta}(z)]^{-1} = [\boldsymbol{\theta}(z)\boldsymbol{\phi}(z)]^{-1} = \sum_{r=0}^{\infty} a_r z^r \ ,$$

tendremos la representación causal para el proceso $\{Y_t\}$, $Y_t - \mu = \sum_{r=0}^{\infty} a_r \varepsilon_{t-r}$, que permite expresar la relación (28) del siguiente modo alternativo

$$\widehat{g}_k = g_k - \sigma^2 \sum_{i=1}^{p+q} \widehat{\lambda}_i a_{k-i} + U_{k,n} \ .$$

Esta expresión puede presentarse de modo matricial con ayuda de la matriz $\mathbf{A}_m$, que tiene la forma

$$\mathbf{A}_m = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_1 & 1 & \cdots & 0 \\ a_2 & a_1 & & 0 \\ \vdots & \vdots & & \vdots \\ a_{(p+q)-1} & a_{(p+q)-2} & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ a_{m-1} & a_{m-2} & \cdots & a_{m-(p+q)} \end{pmatrix} , \tag{37}$$

quedando por tanto

$$\widehat{\mathbf{G}}_m = \mathbf{G}_m - \sigma^2 \mathbf{A}_m \widehat{\boldsymbol{\lambda}} + \mathbf{U}_{m,n} \ , \tag{38}$$

donde $\widehat{\boldsymbol{\lambda}} = (\widehat{\lambda}_1, \widehat{\lambda}_2, ..., \widehat{\lambda}_{p+q})'$ y $\mathbf{U}_{m,n} = (U_{1,n}, U_{2,n}, ..., U_{m,n})'$. De un modo análogo a como se derivan en el enfoque presentado anteriormente, aquí también se pueden obtener las condiciones de ortogonalidad,

$$\mathbf{A}_m^{'} \widehat{\mathbf{G}}_m = O_p(1/n) \ . \tag{39}$$

Como aplicación, sea $m = m_n = \sqrt{n}$ y definamos la matriz de proyección sobre las columnas de $\mathbf{A}_m$, $\Pi_m = \mathbf{A}_m(\mathbf{A}_m^{'}\mathbf{A}_m)^{-1}\mathbf{A}_m^{'}$. Así podríamos de nuevo obtener el estadístico de Box-Pierce (1970):

$$\widehat{Q}_n = n \sum_{k=1}^{m} \widehat{r}_k^2 \overset{D}{\cong} \mathbf{U}_m'(\mathbf{I}_m - \mathbf{\Pi}_m)\mathbf{U}_m \sim \chi^2_{m-(p+q)} \ . \tag{40}$$

APÉNDICE

A.1 Demostración del lema 1

Basta hacer la demostración para cada uno de los términos definidos en el enunciado del propio lema.

- a_k^2. Se tiene $E[a_k^2] = \sum_{t=1}^p \sum_{s=1}^p E(\varepsilon_t \varepsilon_{t+k} \varepsilon_s \varepsilon_{s+k}) = \sum_{t=1}^p E(\varepsilon_t^2 \varepsilon_{t+k}^2) \leq pE(\varepsilon_t^4)$.

- $b_{k,j}^2$. Por los razonamientos usuales, se tiene $E[b_{k,j}^2] =$
$= \sum_{r=0}^\infty \sum_{s=0}^\infty h_r h_s [\frac{1}{n} \sum_{t>p}^{n-k} \sum_{u>p}^{n-k} E(\varepsilon_{t+k} \varepsilon_{t-j-r} \varepsilon_{u+k} \varepsilon_{u-j-s})] \leq \sigma^4 \frac{n-k}{n} \sum_{r=0}^\infty h_r^2 \leq$
$\leq \sigma^4 \sum_{r=0}^\infty h_r^2$.

- $c_{k,j}^2$. Al igual que antes, se tiene $E[c_{k,j}^2] \leq \sigma^4 \sum_{r=0}^\infty h_r^2$.

- d_k^2. Al elevar al cuadrado, se tiene $E[d_k^2] = \frac{1}{n} \sum_{t>p}^{n-k} \sum_{u>p}^{n-k} E(\varepsilon_t \varepsilon_u + \varepsilon_t \varepsilon_{u+k} + \varepsilon_u \varepsilon_{t+k} + \varepsilon_{t+k} \varepsilon_{u+k}) \leq 4\sigma^2$.

- h_{k-j}^2. Los $\{h_r : r \geq 0\}$ están acotados exponencialmente, por tanto $h_r^2 \leq A^2 B^{2r} \leq A^2$,donde $A > 0$ y $0 < B < 1$.

- $e_{k,i,j}^2$. Se tiene $E[e_{k,i,j}^2] \leq E(\varepsilon_t^4)(\sum_{r=0}^\infty |h_r|)^4$ actuando como en $b_{k,j}^2$.

- f_k^2. Es claro que $f_k^2 \leq 1$.

- $l_{k,j}^2$. Se tiene, usando la acotación exponencial para los $\{h_r : r \geq 0\}$,

$$E[l_{k,j}^2] \leq h_{k-j}^2 \sum_{t=n-k+1}^{n} \sum_{s=n-k+1}^{n} E(\varepsilon_t^2 \varepsilon_s^2) \leq E(\varepsilon_t^4) k^2 h_{k-j}^2 \ , \tag{41}$$

donde, usando la acotación exponencial y teniendo en cuenta que $0 < B < 1$, la sucesión de término general $k^2 h_{k-j}^2$ es una sucesión acotada por alguna constante positiva $M > 0$.

Finalmente, como valor de c puede tomarse

$$c = \max\{pE(\varepsilon_t^4),\, \sigma^4 \sum_{r=0}^{\infty} h_r^2,\, 4\sigma^2,\, A^2,\, E(\varepsilon_t^4)(\sum_{r=0}^{\infty} |h_r|)^4,\, 1,\, E(\varepsilon_t^4)M\} \ ,$$

constante positiva universal que no depende de ninguno de los subíndices sobre los que se calcula el supremo. ■

A.2 Caso $ARMA(p,q)$: Desarrollo de Taylor para los residuales

Sea $\delta > 0$ tal que la bola cerrada $\overline{B}(\boldsymbol{\beta},\delta)$ está contenida en el espacio paramétrico causal e invertible asociado al proceso $ARMA(p,q)$ considerado, donde $\boldsymbol{\beta} = (\boldsymbol{\phi},\boldsymbol{\theta})$. Si $t > P$, se tiene el siguiente desarrollo de Taylor, válido si $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| < \delta$:

$$\widehat{\varepsilon}_t = \varepsilon_t(\widehat{\boldsymbol{\phi}}\,, \widehat{\boldsymbol{\theta}},\, \overline{X}_n) = \varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu) + I_t \ , \tag{42}$$

siendo

$$I_t = \sum_{i=1}^{p} D_{t,i}(\widehat{\phi}_i - \phi_i) + \sum_{j=1}^{q} E_{t,j}(\widehat{\theta}_j - \theta_j) + F_t(\overline{X}_n - \mu) + \widehat{Z}_t \ ,$$

donde, $F_t = \partial\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)/\partial\mu = -\sum_{j=0}^{t-1}\pi_j(\boldsymbol{\beta})$ y $\widehat{Z}_t = \sum_{i=1}^{p+q}\sum_{j=1}^{p+q}\partial\varepsilon_t^2(\boldsymbol{\beta}_{t,n},\mu_{t,n})/\partial\beta_i\partial\beta_j$ $(\widehat{\beta}_i - \beta_i)(\widehat{\beta}_j - \beta_j)/2 + \sum_{i=1}^{p+q}\partial\varepsilon_t^2(\boldsymbol{\beta}_{t,n},\mu_{t,n})/\partial\beta_i\partial\mu(\widehat{\beta}_i - \beta_i)(\overline{X}_n - \mu)$. Además,

$$\begin{aligned}
D_{t,i} &= \partial\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)/\partial\phi_i = -\sum_{r=0}^{t-i-1} l_r(X_{t-i-r} - \mu) = -\theta^{-1}(B)(X_{t-i} - \mu) + R_{t-i} \\
E_{t,j} &= \partial\varepsilon_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)/\partial\theta_j = -\sum_{r=0}^{t-j-1} l_r\varepsilon_{t-j-r}(\boldsymbol{\beta},\mu) = -\theta^{-1}(B)\varepsilon_{t-j} + S_{t-j}
\end{aligned}$$

para $i = 1, ..., p$ y $j = 1, ..., q$. Quedando, por tanto, $R_t = \sum_{r=t}^{\infty} l_r(X_{t-r} - \mu)$ y $S_t = \sum_{r=t}^{\infty} l_r\varepsilon_{t-r} + \sum_{r=0}^{t-1} l_r q_{t-r}$, siendo, al igual que anteriormente, $q_t = \sum_{j=t}^{\infty}\pi_j(X_{t-j} - \mu)$. Volviendo a (42), para $t > P$

$$\widehat{\varepsilon}_t = \varepsilon_t - \sum_{i=1}^{p}(\widehat{\phi}_i - \phi_i)\boldsymbol{\theta}^{-1}(B)(X_{t-i} - \mu) - \sum_{j=1}^{q}(\widehat{\theta}_j - \theta_j)\boldsymbol{\theta}^{-1}(B)\varepsilon_{t-j} + \widehat{H}_t \ , \tag{43}$$

donde $\widehat{H}_t = -q_t + \sum_{i=1}^{p}(\widehat{\phi}_i - \phi_i)R_{t-i} + \sum_{j=1}^{q}(\widehat{\theta}_j - \theta_j)S_{t-j} + F_t(\overline{X}_n - \mu) + \widehat{Z}_t$.

A.3 Demostración de las condiciones de ortogonalidad (Proposiciones 1 y 2)

Proposición 1: Caso $AR(p)$

Ya que los estimadores se han obtenido minimizando $u_n(\boldsymbol{\phi})$, los residuales del modelo, por construcción, deben cumplir

$$\left.\frac{\partial u_n(\boldsymbol{\phi})}{\partial \phi_i}\right|_{\phi=\widehat{\phi}} = \sum_{t>p}^{n} \widehat{\varepsilon}_t (X_{t-j} - \overline{X}_n) = 0, \quad j = 1, ..., p. \tag{44}$$

Lo que se pretende ver en esta demostración es que, efectivamente, esta condición implica que

$$\sum_{r=0}^{m-j} h_r \widehat{g}_{r+j} = O_p(\frac{1}{n}), \quad j = 1, ..., p,$$

o lo que es lo mismo, las condiciones de ortogonalidad.

Las ecuaciones de (44) las podemos reescribir como sigue,

$$\frac{1}{n}\sum_{t>p}^{n} \widehat{\varepsilon}_t (X_{t-j} - \overline{X}_n) = \frac{1}{n}\sum_{t>p}^{n} \widehat{\varepsilon}_t (X_{t-j} - \mu) - (\overline{X}_n - \mu)(\frac{1}{n}\sum_{t>p}^{n} \widehat{\varepsilon}_t) = 0 \ .$$

El término que resta multiplicado por n, $(\overline{X}_n - \mu)(\sum_{t>p}^{n} \widehat{\varepsilon}_t) = \sqrt{n}(\overline{X}_n - \mu)\sqrt{n}$ $(n^{-1}\sum_{t>p}^{n} \widehat{\varepsilon}_t)$, será $O_p(1)$ en virtud del resultado (6) del teorema 1 y el teorema central del límite. Por tanto, el primer término es también $O_p(n^{-1})$.

Considerando ahora la representación causal del proceso, podemos escribir $\sum_{t>p}^{n} \widehat{\varepsilon}_t (\sum_{r=0}^{m-j} h_r \varepsilon_{t-j-r})/n = n^{-1}\sum_{t>p}^{n} \widehat{\varepsilon}_t (X_{t-j} - \mu) - \sum_{t>p}^{n} \widehat{\varepsilon}_t (n^{-1}\sum_{r>m-j}^{\infty} h_r \varepsilon_{t-j-r})$. El primer sumando es, como acabamos de ver, $O_p(n^{-1})$. Por otro lado, para tratar de acotar el segundo, podemos usar la expresión (3) y con ella, reescribir el segundo sumando como sigue, $\sum_{t>p}^{n} \widehat{\varepsilon}_t (\sum_{r>m-j}^{\infty} h_r \varepsilon_{t-j-r}) = \sum_{t>p}^{n} \varepsilon_t q_{t,j} - \sum_{i=1}^{p} (\widehat{\phi}_i - \phi_i)[\sum_{t>p}^{n} (X_{t-i} - \mu) q_{t,j}] + Q_n \sum_{t>p}^{n} q_{t,j}$ donde $q_{t,j} = (\sum_{r>m-j}^{\infty} h_r \varepsilon_{t-j-r})$ y $Q_n = -\widehat{\phi}(1)(\overline{X}_n - \mu)$. Si acotamos cada uno de estos sumandos veremos que los dos primeros son $o_p(1)$ mientras que el tercero es $O_p(n^{-1})$. Por tanto, se concluye que $\sum_{t>p}^{n} \widehat{\varepsilon}_t (\sum_{r>m-j}^{\infty} h_r \varepsilon_{t-j-r}) = O_p(n^{-1})$.

Si ahora dividimos este sumatorio en dos partes y, de nuevo, usamos la relación entre $\widehat{\varepsilon}_t$ y ε_t que se describe en (3) podemos llegar a escribir,

$$O_p(\frac{1}{n}) = \sum_{r=0}^{m-j} h_r(\frac{1}{n}\sum_{t>p}^{p+(j+r)} \widehat{\varepsilon}_t\varepsilon_{t-j-r}) + \sum_{r=0}^{m-j} h_r(\frac{1}{n}\sum_{t>p}^{n-(j+r)} \varepsilon_t\widehat{\varepsilon}_{t+(j+r)}) = \sum_{r=0}^{m-j} h_r\widehat{g}_{r+j} +$$

$$+\sum_{r=0}^{m-j} h_r(\frac{1}{n}\sum_{t>p}^{p+(j+r)} \widehat{\varepsilon}_t\varepsilon_{t-j-r}) + \sum_{r=0}^{m-j} h_r(\frac{1}{n}\sum_{t>p}^{n-(j+r)} [\sum_{i=1}^{p}(\widehat{\phi}_i - \phi_i)(X_{t-i} - \mu) - Q_n]\widehat{\varepsilon}_{t+(j+r)}) \, ,$$

y llegados a este punto sólo quedaría por mostrar que el segundo y tercer sumando son $O_p(n^{-1})$. Para comprobar ambas acotaciones es aconsejable usar de nuevo la expresión (3). ■

Proposición 2: Caso $ARMA(p,q)$

Se procede de un modo similar a la demostración del caso $AR(p)$ (Proposición 1) pero teniendo en cuenta que las condiciones de ortogonalidad exacta de las que se parte son dos en este caso y se definen, análogamente al caso $AR(p)$, con las expresiones,

$$\left.\frac{\partial u_n(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial\phi_j}\right|_{(\boldsymbol{\phi},\boldsymbol{\theta})=(\widehat{\boldsymbol{\phi}},\widehat{\boldsymbol{\theta}})} = \sum_{t>P}^{n}\widehat{\varepsilon}_t\partial\widehat{\varepsilon}_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)/\partial\phi_j = \sum_{t>P}^{n}\widehat{\varepsilon}_t\widehat{D}_{t,j} = 0, \quad j = 1,...,p \, ,$$

y

$$\left.\frac{\partial u_n(\boldsymbol{\phi},\boldsymbol{\theta})}{\partial\theta_h}\right|_{(\boldsymbol{\phi},\boldsymbol{\theta})=(\widehat{\boldsymbol{\phi}},\widehat{\boldsymbol{\theta}})} = \sum_{t>P}^{n}\widehat{\varepsilon}_t\partial\widehat{\varepsilon}_t(\boldsymbol{\phi},\boldsymbol{\theta},\mu)/\partial\theta_h = \sum_{t>P}^{n}\widehat{\varepsilon}_t\widehat{E}_{t,h} = 0, \quad h = 1,...,q \; .$$

La primera de las cuales se utiliza para mostrar una parte del producto matricial, en concreto $\sum_{r=0}^{m-j} h_r\widehat{g}_{r+j} = O_p(\frac{1}{n})$, $j = 1, ..., p$, mientras que, a su vez, la segunda implica $\sum_{s=0}^{m-h} l_s\widehat{g}_{s+h} = O_p(\frac{1}{n})$, $h = 1, ..., q$. ■

REFERENCIAS

1. Box, G.E.P. & Pierce, D. A. (1970): "Distribution of residual autocorrelation in autorregressive integrated moving average time series models", *Journal of the American Statistical Association, 65, 332*

2. Brockwell, P. J. & Davies, R. A. (1991): *"Time Series Analysis: Forecasting and Control".* Holden-Day, S. Francisco, 2nd. ed.